

Infaux ou Vérité?

By Les Reptiliens



AMAR James
BEN-YOUNES Lina
PEYREDIEU DU CHARLAT Blanche
VERON Clément
VIGIER Albane

<https://gitlab-ovh-02.cloud.centralesupelec.fr/clement.veron/infaux-et-verite/-/tree/main>





Description de l'application

- Estimer la **probabilité de crédibilité** d'un tweet en s'appuyant sur :
 - son **contenu** (*texte et pseudo de l'utilisateur*)
 - les **relations** de l'utilisateur (*si l'utilisateur est relié à des comptes considérés comme fiables ou pas*)



- A partir de ce calcul, estimer la **proportion de fake news** parmi un ensemble de tweets portant sur un **événement donné**.
- **Cadre** : politique en France



Analyse des besoins

- Outil d'**analyse de l'opinion publique** et de la **qualité de l'information**.
- **User Stories:**
 - *En tant que média, je veux me renseigner sur les nouvelles circulant sur Internet afin de mieux informer.*
 - *En tant qu'organisateur d'événement / personnalité / marque, je veux observer les informations qui sont échangées à mon propos pour améliorer ma communication.*








Description des MVPs

Nous avons divisé le MVP en **3 objectifs** :

- Analyse **Textuelle** (*MVP1*)
- Analyse **Sociale** (*MVP2*)
- Analyse de l'**Événement** (*MVP3*)

Chaque MVP renvoie un nombre **entre 0 et 1**

Un objectif = un package sur le dépôt Gitlab

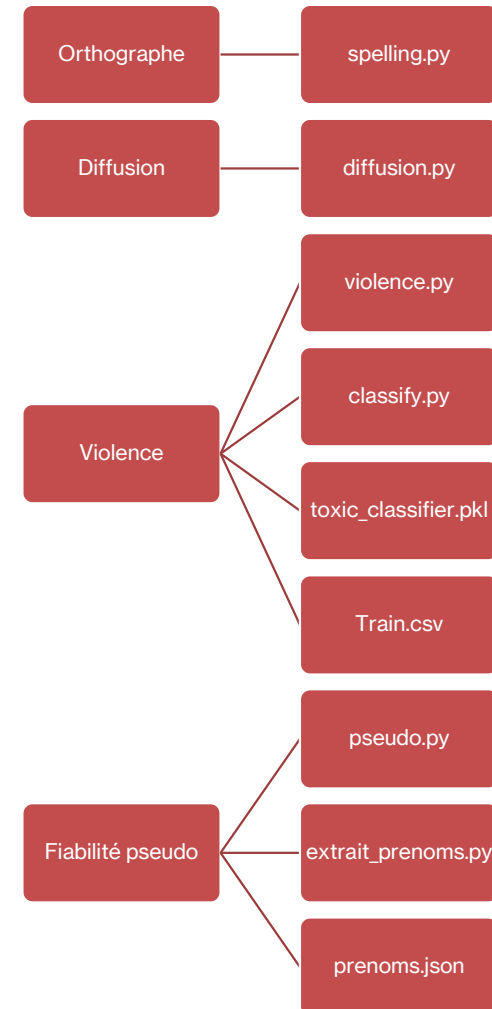
Name
 affichage
 news_analysis
 social_analysis
 textual_analysis
 twitter_collect
 README.md
 infaux_ou_verite.py

MVP1 : Textual Analysis

- **En entrée :** un tweet = texte + nom d'utilisateur (*screen_name*)
- **Fonctions :**
 - **Orthographe** (James)
 - **Diffusion** (Blanche et Albane)
 - **Violence** (Lina)
 - **Fiabilité Pseudo** (Clément)
 - **Fonction générale** (Blanche)

Cf : Fausses Nouvelles : guide des questions à se poser face à une information, site du Gouvernement
(<https://www.gouvernement.fr/fausses-nouvelles-guide-des-questions-a-se-poser-face-a-une-information>)

Fake News: 10 astuces pour les repérer, Centre de Culture Scientifique Technique et Industrielle de Mines Saint-Étienne
(<https://www.larotonde-sciences.com/jeux-et-experimentations/fake-news/>)





Démonstration MVP1

```
File Edit Selection View Go Run Terminal Help
mvp1.py - infaux-et-verite - Visual Studio Code

EXPLORER
infaux-et-verite
  > __pycache__
  contacts.py
  followers.py
  identification.py
  mvp2.py
  replying.py
  retweet.py
  textual_analysis
    > __pycache__
    .gitkeep
    classify.py
    diffusion.py
    extrait_prenoms.py
    mvp1.py
    prenom.json
    pseudo.py
    spelling.py
    toxic_classifier.pkl
    train.csv
    violence.py
    twitter_collect
    infaux_ou_verite.py
  README.md
  tweets.json

textual_analysis > mvp1.py
1 from textual_analysis.diffusion import Diffusion
2 from textual_analysis.violence import Violence
3 from textual_analysis.pseudo import prenom_pseudo
4 from textual_analysis.spelling import misspelling_average
5
6
7 def generale1(tweet, screen_name):
8     """prend en argument un tweet sous forme de texte et le nom d'utilisateur
9     et renvoie la probabilité que ce soit une fakenews, d'après une analyse textuelle"""
10    s = 0
11    s += Diffusion(tweet)*0.2
12    s += misspelling_average(tweet)*0.4
13    s += prenom_pseudo(screen_name)*0.1
14    s += Violence(tweet)*0.3
15    return s
16 #print (generale1("jeanlassalle", "Petite surprise à 15h."))

PROBLEMS OUTPUT TERMINAL DEBUG CONSOLE
linab@LAPTOP-OHLKG40C MINGW64 ~/Documents/coding weeks/infaux-et-verite (main)
$ python infaux_ou_verite.py
AT est sélectionné

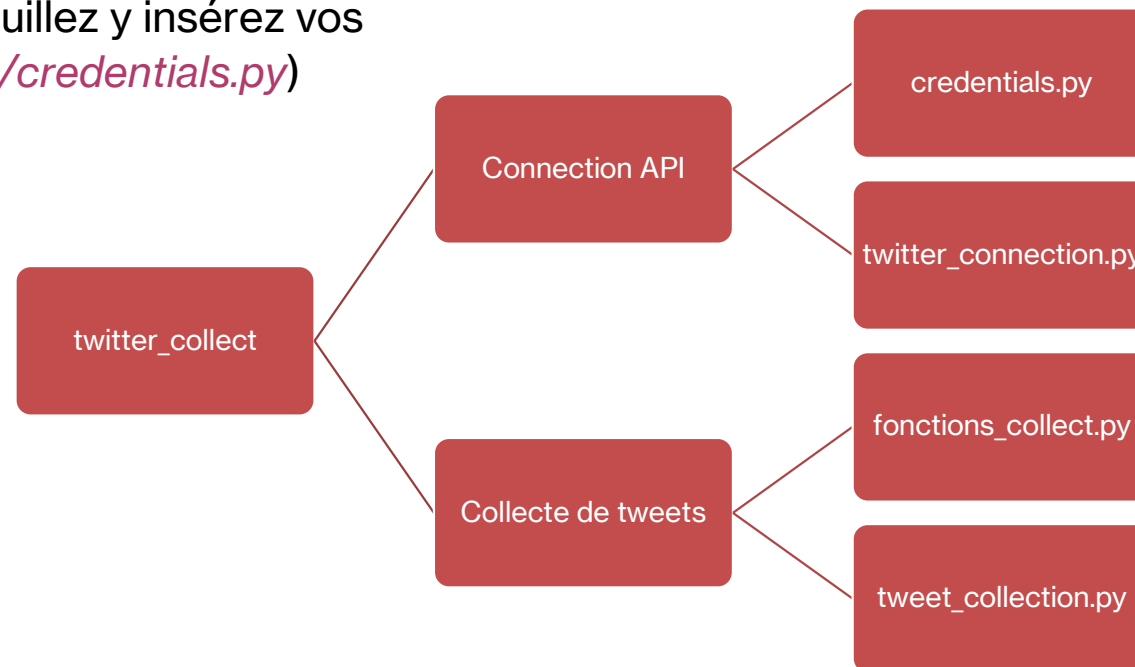
linab@LAPTOP-OHLKG40C MINGW64 ~/Documents/coding weeks/infaux-et-verite (main)
$ python infaux_ou_verite.py
AT est sélectionné

linab@LAPTOP-OHLKG40C MINGW64 ~/Documents/coding weeks/infaux-et-verite (main)
$
```



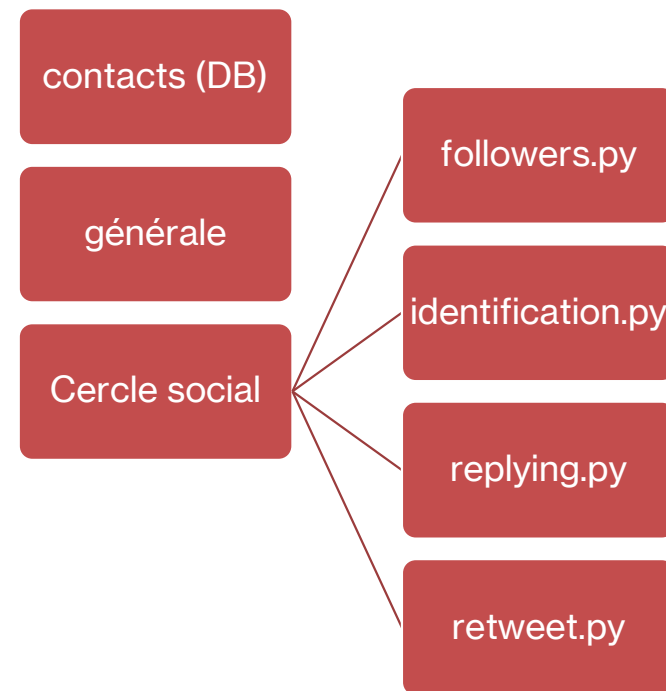
Collecte des tweets

- Package **twitter_collect** contenant des fonctions plus générale utiles dans les MVP2 et MVP3
- Pour utiliser l'application, veuillez y insérez vos **credentials**. ([*twitter_collect/credentials.py*](#))



MVP2 : Social Analysis

- **En entrée :** nom d'utilisateur (*screen_name*)
- **DataBases:** (Blanche et Clément)
 - Comptes fiables : *user_id*, *screen_name*
 - Comptes non fiables : *user_id*, *screen_name*
- **Fonctions :**
 - Réponses (James)
 - Retweets (Clément)
 - Identifications (Albane et Lina)
 - Followers (Clément)
 - **Fonction générale :** note sur la fiabilité du cercle de l'utilisateur (Lina)





Démonstration MVP2

```
File Edit Selection View Go Run Terminal Help
infaux_ou_verite.py - infaux-et-verite - Visual Studio Code

EXPLORER
  INFAUX-ET-VERITE
    > affichage
    > social_analysis
      replying.py
      infaux_ou_verite.py

1 from affichage.fenetre import main_fenetre
2 from textual_analysis.mvp1 import generale1
3 from social_analysis.mvp2 import generale2
4 from news_analysis.mvp3 import generale3
5 from twitter_collect.fonctions_collect import collect_by_user, screen_name_to_user_id
6
7 if __name__ == '__main__':
8     main_fenetre(generale1, generale2, generale3)
9     # print(len(collect_by_user(screen_name_to_user_id('JoyTaniguchi'))))
10    # print(generale3(['Cheveux', 'cheval']))
11

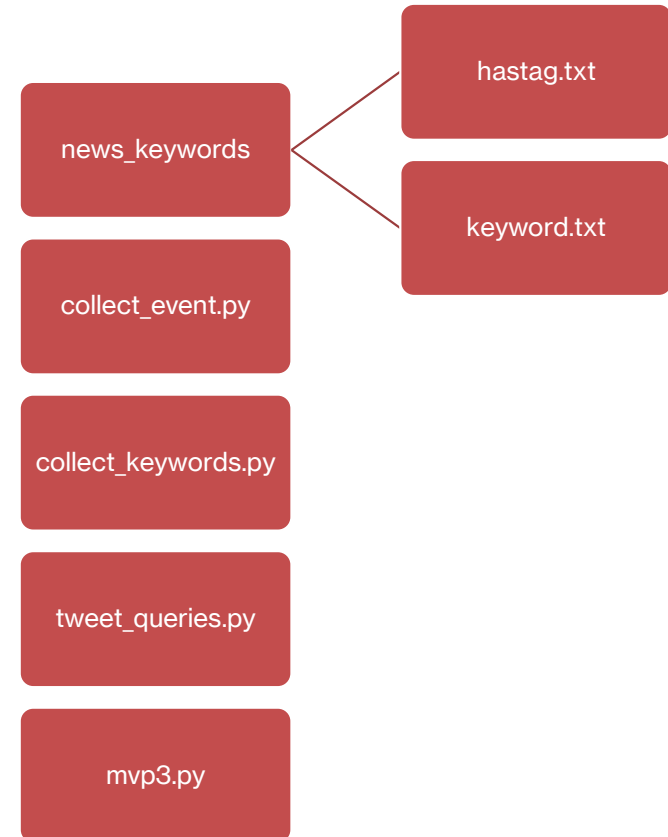
PROBLEMS OUTPUT TERMINAL DEBUG CONSOLE
Enumerating objects: 47, done.
Counting objects: 100% (47/47), done.
Delta compression using up to 4 threads
Compressing objects: 100% (37/37), done.
Writing objects: 100% (37/37), 16.81 KiB | 400.00 KiB/s, done.
Total 37 (delta 9), reused 0 (delta 0), pack-reused 0
To gitlab-ovh-02.cloud.centralesupelec.fr:clement.veron/infaux-et-verite.git
f7c0f26..11e73f9 main -> main

linab@LAPTOP-OHLKG40C MINGW64 ~/Documents/coding weeks/infaux-et-verite (main)
$
```

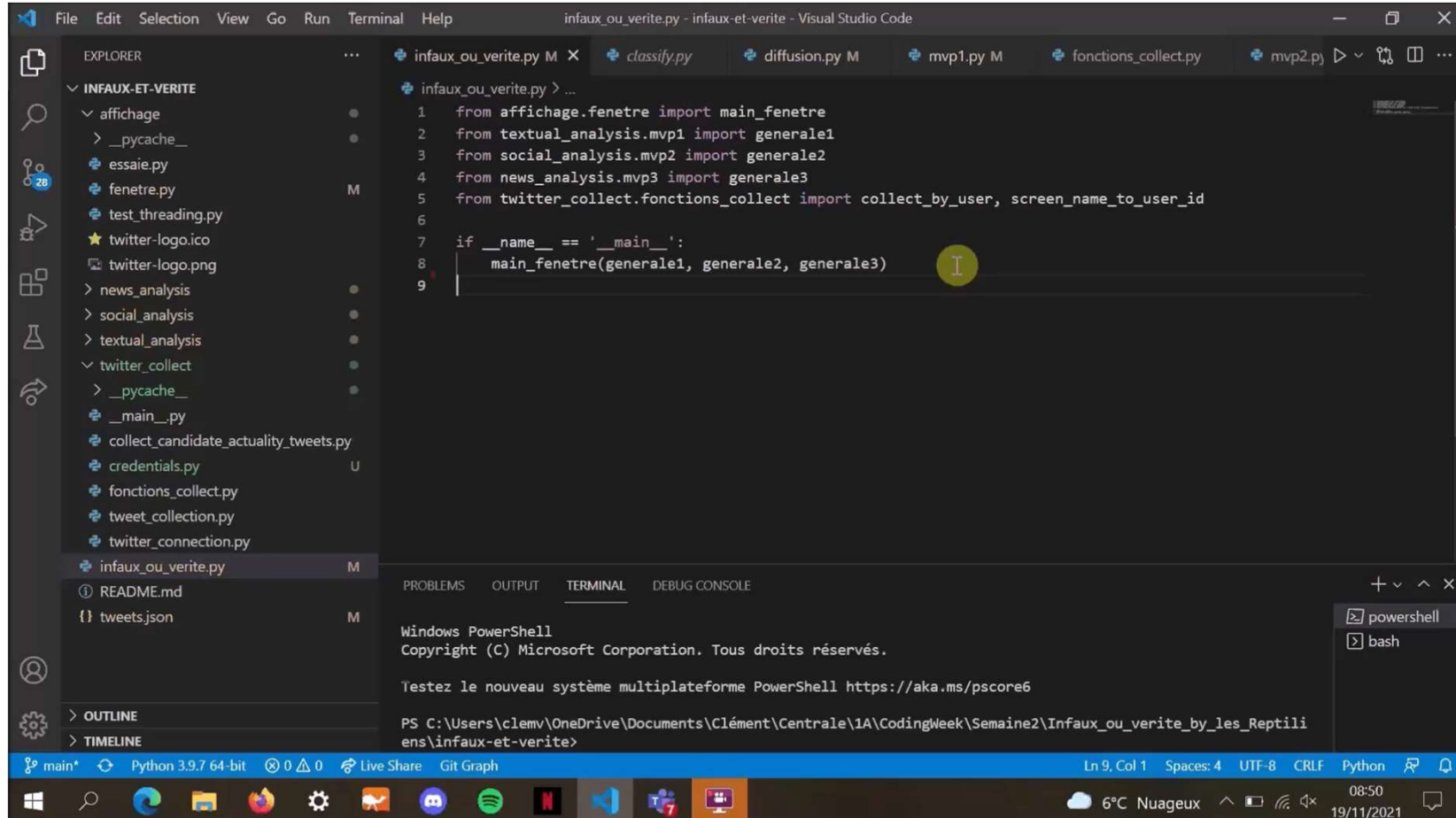


MVP3 : News Analysis

- **En entrée:** mots-clés et hashtags qui décrivent l'événement
- **En sortie:** estimation de la **proportion de fake news** parmi les tweets à propos de cet événement
- **Fonctions:**
 - **Collection des mots-clés** (*fichier txt*) (Blanche)
 - **Collection des tweets** contenant les mots-clés (Albane)
 - **Générale** : analyse des tweets et calcul de proportion (**MVP1 + MVP2**) (James)



Démonstration MVP3



```
1 from affichage.fenetre import main_fenetre
2 from textual_analysis.mvp1 import generale1
3 from social_analysis.mvp2 import generale2
4 from news_analysis.mvp3 import generale3
5 from twitter_collect.fonctions_collect import collect_by_user, screen_name_to_user_id
6
7 if __name__ == '__main__':
8     main_fenetre(generale1, generale2, generale3)
9
```


Windows PowerShell
Copyright (C) Microsoft Corporation. Tous droits réservés.

Testez le nouveau système multiplateforme PowerShell <https://aka.ms/pscore6>

PS C:\Users\clemv\OneDrive\Documents\Clément\Centrale\1A\CodingWeek\Semaine2\Infaux_ou_verite_by_les_Reptili
ens\infaux-et-verite>




Affichage (Clément)

 Infaux ou vérité

Que cherchez-vous ?

☒ Analyse textuelle ☐ Analyse sociale ☐ Événement

Valider

 Infaux ou vérité

Permet d'obtenir la probabilité que le tweet soit une fake news par analyse textuelle

Texte du tweet

Deja ce n'est pas "oune" mais "un" un peut d'effort pour l'orthographe en roumain ça fait mal a cœur pour nous les roumain de voir des fautes d'orthographe aussi bête mais c'est pas grave

Nom d'utilisateur

Ncl78

Valider

Score de non fiabilité :

0.34411764705882353

Retour



Conclusion

Limites du modèle :

- Les bases de données d'utilisateurs fiables et non fiables sont très limitées.
- L'analyse des champs lexicaux avec Textblob se fait à partir d'une base de donnée en anglais, il faut donc traduire les tweets français pour les analyser.
- L'analyse de l'orthographe n'est pas faite dans le contexte de la phrase (accords, conjugaison, etc.).
- Les poids des critères sont choisis arbitrairement.
- On n'a accès qu'à un nombre limité de tweet par minute, il est donc impossible de véritablement établir les liens entre un utilisateur et son cercle (ambition du MVP2)
- Le traitement des tweets devient trop lourd si on prend en compte plus d'un tweet par personne
- Problème de mémoire RAM sur le module violence

Idée de solutions :

- Enrichir la base de données des utilisateurs fiables / non fiables lorsque l'application analyse un l'utilisateur
- Il faudrait réussir à se procurer des bases de données en français
- Il faudrait utiliser un module qui analyse plus précisément la grammaire et l'orthographe (jampell).
- On pourrait proposer à l'utilisateur de l'application de déterminer lui même le poids qu'il souhaite attribuer à chaque critère